International Journal of Innovative Research in Science and Technology
Volume 01, Issue 01, August 2021
www.ijirst.com

Evaluation of Deep Learning Methods in Twitter Statistics Emotion Evaluation

Saidireddy Maligireddy[1], Dr. M. V. P. Chandra Sekhara Rao[2]
[1](Research Scholar, ANU College of Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India)
[2](Professor, Dept. of CSE, R.V.R.&J.C College of Engineering, Acharya Nagarjuna University (ANU), Guntur, Andhra Pradesh, India)

To Cite this Article
Saidireddy Maligireddy, Dr. M. V. P. Chandra Sekhara Rao, "Evaluation of Deep Learning Methods in Twitter Statistics Emotion Evaluation", International Journal of Innovative Research in Science and Technology, Vol. 01, Issue 01, August 2021, pp.:001-007.

Abstract: *This analysis compares and contrasts a variety of methods for assessing emotions in Twitter data. Deep learning (DL) methods have gained momentum in this field among academics, who collaborate on a level playing field to tackle a wide variety of problems. CNNs, which are used to locate pictures, and recurrent neural networks (RNNs), which may be utilized successfully in natural language processing (NLP), are two types of neural networks. For this reason, two types of neural networks are explicitly utilized. These images are used to assess and compare CNN ensembles and variants, as well as RNN category networks with long-term memory (LSTM). We also associate clothing with the type phrase embedding structures Word2Vec and the global phrase representation vectors (Glove). To put these methods to the test, we utilized information from the Seminal (Seminal), one of the most well-known international workshops on the internet. Different trials and combinations are used, and the better results for each variation are linked to their average efficiency. This study adds to the area of sentiment analysis by assessing the outcomes, benefits, and drawbacks of various methods using an evaluation method that use a single testing system for the same dataset and machine configuration.*

Keywords*: Emotion estimation, in-depth learning, neural network convolution, LSTM, phrase embedding models, and Twitter statistics*

I. Introduction

Emotion assessment has been acknowledged by a wide range of people with various interests and motivations as a result of the rise in the use of social media in recent years. Extracting information from these papers is becoming more important as people across the globe are able to express their views on approximately specified topics related to government, education, travel, subculture, commercial goods, and well-known concerns. Understanding customers' emotions as they express themselves via their communications in different formats has shown to be important information for assessing people's perceptions of a specific problem, in addition to data connected to visited locations, buying choices, and so on. The classification of a text's polarity in consumer pride, disappointment, or neutrality terms is a common approach. Polarity may vary from effective to poor in terms of grading or a wide range of levels, but it typically refers to textual material emotions that range from joyful to sad. For the extraction of appropriate functions and categorization of texts into applicable polarity markings, a variety of techniques are employed, with an emphasis on one-of-a-kind herbal language processing and system learning approaches. Despite the fact that deep learning methods have been popular for many years, many deep neural networks have been successfully deployed on the ground. In terms of sentiment analysis, neural networks and LSTM networks, in particular, have shown to be effective. Their effectiveness has been shown in a variety of empirical studies, both alone and in combination. In the area of natural language processing, most methods for extracting features from words, such as Word2Vec and global phrase representation vectors (Glove), are popular. The aforementioned methods' accuracy is good, but not great, which is why sentiment analysis remains a continuous and accessible research issue. These scholars want to expand or enhance current techniques. Because current methods include a wide range of network design, tuning, and other functions, a study of the tactics that have

previously been employed is necessary to understand their limits and the complexity of sentiment assessment. This article adds to the field by comparing the most popular deep mastering methods and configurations inside a single test system, all of which are based on an agreed data source that is based on Twitter data. The following sections make up the document:

The following work in this region is included in Phase 2. Section 3 explains how to use the neural network and some of the unique variants that may be used. Segment 4 discusses the consequences, compares and contrasts the best methods, and explains the results. Section 5 brings the article to a close.

II. Historical Context

The assessment of emotions and opinion mining have been a focus for academics all over the world, thanks to the growth and popularity of social media and a variety of structures that enable people to express their views on a variety of topics. The authors listed the various methods used up to that point in time in a 2008 paper. In recent years, deep neural networks have proven to be especially effective in emotion evaluation tasks. Because CNN fully responds to the dimension discount problem and LSTM networks perform a category of RNN with transient or sequential data, neural networks and recurrent neural networks have been widely used. CNN architectures may be utilized with output for sentence class in the developers' breakthrough work. In contrast, it was shown that although CNN performed slightly better than traditional methods, the RNN's performance When CNN and LSTM networks were combined, it outperformed state-of-the-art techniques and offered a significant advantage. For comparable findings around the same time, GRU networks, which were developed in 2014, may be used instead of LSTM. According to a study of deep learning methods in emotion analysis, phrase embedding is done directly using two tools, Word2Vec or Glove thesis days. In every nation on the globe, Twitter is one of the most extensively utilized social networking sites. As a consequence, the media must be able to extract public opinion from a wide range of subjects from tweets, evaluate the impact of specific events, and distinguish between emotions. The early work of emotional analysis relied entirely on bi-grams, unigrams, particular polarity functions, and devices that were familiar with classifiers like Bayesian networks or vector supporting machines for extracting power. In the years that followed, other scientific contests were organized in odd locations throughout the world in order to appeal to the interests of researchers. For the last 13 years, the International Semantic Evaluation Workshop has held competitions in this area.

A. Thorough understanding of techniques carries a lot of weight nowadays. The linked research aims to use especially unusual combinations of neural networks and different implementations of word embedding functions to get an advantage over the competition. The total results of the Twitter knowledge sentiment assessment were used to define many researches. The authors suggest employing two kinds of word embedding, Word2Vec and Glove, in two separate CNN settings, with the results pooled in a random forest grouping. In every other research, the authors use embedding systems that have been trained in lexical, element-of-speech, and emotion embedding, all of which may be utilized to kick off a deep CNN framework. The authors presented two LSTM networks that are mostly bidirectional. The word embedding is done using a glove. Others, as said, use a combination of CNN and LSTM networks. When comparing Word2Vec, Glove, and Fast Text, the authors discovered that Glove generated negative outcomes when compared to the other two. RCNNs and CNNs were ultimately combined in an efficient way. Regardless of the findings of the aforementioned study, comparing and assessing the location of a data collection, network architecture, or a specific configuration and tuning became very challenging. This issue prompted this study, which aimed to construct a distinct structure with the goal of comparing different techniques and explaining the advantages and disadvantages of each unique arrangement. Technology, technology, technology the dataset, phrase integration models, and their configurations, as well as the one-of-a-kind deep neural network configurations used in this section, are all described in this section. Because the findings for LSTM networks and CNNs are comparable, GRU networks and RCNNs are not included in the following settings. A. Dataset & Preprocessing In Seme Val competitions, three data sets are used to completely execute a series of different data sets. We used the whole SemEval2014 task9-SubTaskB statistics, as well as the entire SemEval2015 knowledge Task4 and SemEval2017 progress data, which totaled 32,000 tweets. The next stage is to collect tweets with a total of 662,000 words and a vocabulary of approximately 10,000 keywords to improve the application's overall efficiency during the educational process. As a result, an additional preprocessing operation was carried out to delete and change certain characters. Turning all letters into situations, eliminating a few special characters and emoticons, and marking URLs were all part of this job.

B. Word2Vec and Glove are used to embed the expression embedding templates provided with this look in Word. The Word2Vec model was converted into a 25-dimensional word vector centered on the previously reported dataset. The Word2Vec setup was complete with the CBOW model. Frequently, words that occurred fewer than five

times were removed. Finally, the most effective sentences were cut down to five words. GloVe's expression vectors have been utilized. They're 25-dimensional vectors containing 2 billion tweets, which is a far larger dataset than the SemVal details-derived data set. Apply the following equation to normalize all vectors.

$$v_i' = \frac{v_i - v_{min}}{v_{max} - v_{min}} \qquad (1)$$

Where is the standardized, I value of the vector's 25-dimensional minimum and maximum value?

1) Sentencing vectors When the twitter word vectors are merged to create a unique vector, sentence vectors are created. After experimenting with different lengths, we came up with 40-word sentences. Because the length of tweets fluctuates, any additional words in a tweet have been removed. When the terms in a tweet were less than 40 characters long, they were repeated until the necessary length was reached. Zero padding is another method for filling in missing words in a sentence. This approach used zero padding only for words that were not in the lexicon.

2) A Sentence's Sections During the prediction phase, one option to word embedding is to split the word vectors into regions in order to keep information in a single phrase and long-distance dependence throughout the sentence. Punctuation marks are used to divide words in a sentence. Each area in the current layout has 10 words, and each phrase has eight areas. If any words or regions are missed, null padding is used to fill in the gaps. The sequence of regions in a phrase is shown in Figure 1.

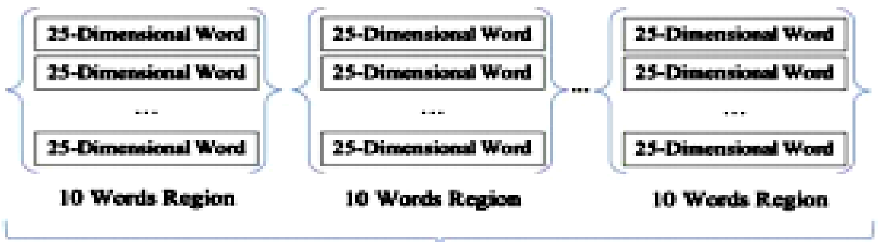

Figure no 1: Regional sentence form. Each sentence consists of eight regions, and each area consists of 10 25-dimensional terms. Null padding shall be added in the case of missed terms or regions to cover the missing regions

Finally, the dataset is divided into two parts, one of which is non-regional and the other of which is regional. The input scale is 1000 in the first case (a sentence includes 40 terms, each of which is 25), and 2000 in the second case (a phrase contains eight regions, each of which contains ten words of size 25). A kind of computer network is a neural network. CNN and LSTM networks are suggested neural network architectures for evaluating twitter data. An SVM classifier is also used in one instance. Datasets from both regional and non-regional regions were used to evaluate both networks. There were eight different network setups proposed. As previously stated, RCNN and GRU networks are not used since they were unable to compete with CNN and LSTM networks in our testing. Sigmoid activation and 300 epochs were used to train all of the networks.

1) Only one CNN channel There is just one 1-dimensional CNN layer in this network. This structure is shown in Figure 2, in which the phrase vector is divided into 12 kernels with sizes ranging from one to three. (In comparison to other kernel variants, it performed better in our testing). 1 to 3 is the maximum pooling layer height. For the following CNN configurations, the CNN parameters would be the same. Finally, polarity causes a 3-dimensional construct to react in one of three ways: positively, negatively, or neutrally.

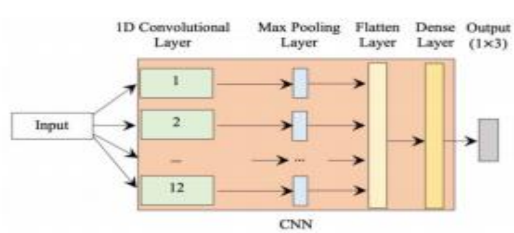

Figure no 2: CNN setup for positive, neutral and negative polarity estimation with a single layer and a 3-dimensional performance

2) A single-node LSTM network A single LSTM layer with a 20% drop-off is used in this setup. The performance varies from 1 to 3 for predicting polarity. (either positively, neutrally, or negatively).

3) The purpose of this arrangement is to aggregate the outputs of the individual CNN and LSTM networks for analysis. A soft vote based on network outputs determines the prediction response. Figure 3 shows a setup in which the CNN and LSTM modules are constructed to the same degree as the previous two configurations (for CNN 12 kernels of size 1 to 3 and an overall pooling layer of 1 to 3).

4) A single 3-layer CNN and LSTM network In this configuration, a one-dimensional, three-layer CNN and a single LSTM network layer are utilized. In this setup, the input is routed into a three-layer CNN (see Figure 4). The input is 1000 when concentrating on words (non-regional) and 2000 when focused on regions. (regional).

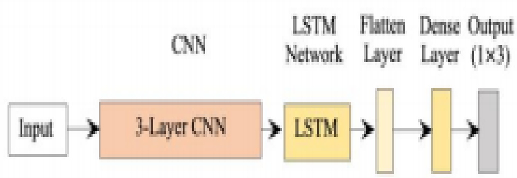

Figure no 3: 3-layer CNN and LSTM network hybrid

5) Various types of CNN and LSTM networks In the new architecture, the data is split into specific components, non-regional input words, and regional input areas. These elements are fed into each CNN as input. The performance of each CNN is then sent into a single LSTM network as an input. Figure 5 depicts the network configuration. We have 40 or eight CNNs, depending on the kind of input (40 words or eight regions). Each CNN network uses 12 kernels, as previously stated.

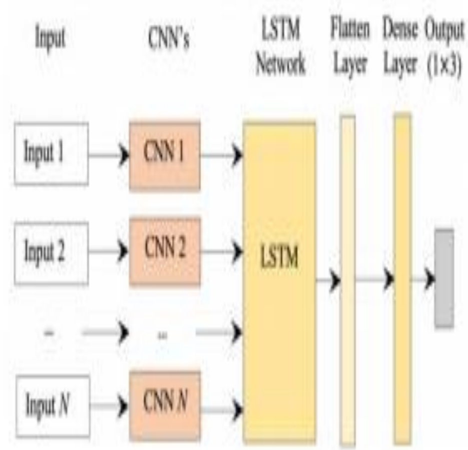

Figure no 4: CNN and LSTM networks with an input separated into N inputs. Mix. N is equivalent to 40, if the input is non-regional, or 8, if the input is regional

6) CNN single 3-layer LSTM bidirectional network This setup is similar to (5), except this time bidirectional LSTM networks are used. The goal of this setup is to evaluate two-way LSTM network performance to basic LSTM network performance. 7) A CNN-based bidirectional LSTM network. This setup is identical to (6), except instead of bidirectional LSTM networks, bidirectional LSTM networks are used.

III. Result

The accuracy, precision recall, and F-measure (F1) output results from prior networks are shown in this section. setups.as defined in the following equations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F\_Measure = \frac{2*Recall*Precision}{Recall+PRecision} \qquad (5)$$

In the case above, the genuine negative, false positive, and false negative predictions are the real positive forecasts. Tables I and II demonstrate the efficiency results of the proposed Wor2Vec and GloVE word integration techniques utilizing CNN and LSTM networks, respectively. To begin, the GloVe gadget increases the efficiency of almost all setups (5 percent -7 percent). As a result, compared to GloVe-based pretrained word vectors with a considerably larger training dataset, Word2Vec has achieved word vectorization using a very small training dataset of approximately 32,000 tweets. The second point is that the device improves efficiency by utilizing numerous CNNs with LSTM networks instead of simple setups, regardless of the term embedding method (3 percent -6 percent). We'll find that, in contrast to other choices, certain settings nearly always provide the best results. Last but not least, splitting text elements into regions does not substantially enhance setup efficiency in most cases (1 percent -2 percent). When SVM classification is employed instead of a soft-voting method, the result may be somewhat poorer. Finally, utilizing two-way LSTM networks instead of basic LSTM networks offers little advantage, as shown by the excellent quality of the results (the structure of words in a sentence). The strongest findings of this research are compared to prior studies that utilized comparable neural networks in Table III. The current research shows a comparable but slightly lower literary efficiency than the prior one, as can be shown (6 percent difference). This is to be anticipated, and it may be ascribed to the different databases and unique techniques used to shape the dataset or modify the network in prior research. Furthermore, rather than obtaining the best possible results in accordance with prior trials, the emphasis of this research was on assessing and comparing the various deep neural networks and word embedding systems within a specific context. The best accuracy result in the literature (65 percent) is unsatisfactory at this stage, showing that deep learning techniques for sentiment analysis are far from guaranteeing an output comparable to other sectors where the same networks are applied with more success (e.g., deep learning networks for object recognition in images).

Table no 1: Predictions of various CNN and LSTM versions of the Word2Vec word embedding method for non-regional and regional settings based on a sample of roughly 32,000 tweets

Network Model	Type	Recall	Prec.	F1	Acc.
1. Single CNN network	N-R[a]	0.33	0.35	0.33	0.49
	R	0.32	0.34	0.33	0.51
2. Single LSTM network	N-R	0.43	0.51	0.39	0.51
	R	0.44	0.49	0.39	0.50
3. Individual CNN and LSTM Networks	N-R	0.43	0.47	0.37	0.50
	R	0.46	0.52	0.42	0.52
4. Individual CNN and LSTM Networks with SVM classifier	N-R	0.45	0.46	0.43	0.49
	R	0.42	0.54	0.38	0.51
5. Single 3-Layer CNN and LSTM Networks	N-R	0.41	0.52	0.40	0.46
	R	0.40	0.46	0.35	0.48
6. Multiple CNN's and LSTM Networks	N-R	0.43	0.47	0.37	0.50
	R	0.46	0.52	0.43	0.52
7. Single 3-Layer CNN and bi-LSTM Networks	N-R	0.42	0.45	0.39	0.48
	R	0.42	0.47	0.36	0.48
8. Multiple CNN's and bi-LSTM Networks	N-R	0.43	0.50	0.38	0.51
	R	0.46	0.51	0.44	**0.52**

Embedding Word System: Word2Vec					
Network Model	Type	Recall	Prec.	F1	Acc.
1. Single CNN network	N-R[a]	0.33	0.35	0.33	0.49
	R	0.32	0.34	0.33	0.51
2. Single LSTM network	N-R	0.43	0.51	0.39	0.51
	R	0.44	0.49	0.39	0.50
3. Individual CNN and LSTM Networks	N-R	0.43	0.47	0.37	0.50
	R	0.46	0.52	0.42	0.52
4. Individual CNN and LSTM Networks with SVM classifier	N-R	0.45	0.46	0.43	0.49
	R	0.42	0.54	0.38	0.51
5. Single 3-Layer CNN and LSTM Networks	N-R	0.41	0.52	0.40	0.46
	R	0.40	0.46	0.35	0.48
6. Multiple CNN's and LSTM Networks	N-R	0.43	0.47	0.37	0.50
	R	0.46	0.52	0.43	0.52
7. Single 3-Layer CNN and bi-LSTM Networks	N-R	0.42	0.45	0.39	0.48
	R	0.42	0.47	0.36	0.48
8. Multiple CNN's and bi-LSTM Networks	N-R	0.43	0.50	0.38	0.51
	R	0.46	0.51	0.44	**0.52**

Table no 2: Prediction of various CNN and LSTM network variants from approximately 32,000 tweets using the GloVe word embedding device in non-regional and regional setups

Network Model	Type	Recall	Prec.	F1	Acc.
1. Single CNN network	N-R[a]	0.44	0.41	0.4	0.54
	R	0.35	0.31	0.31	0.48
2. Single LSTM network	N-R	0.5	0.58	0.48	0.55
	R	0.51	0.55	0.51	0.55
3. Individual CNN and LSTM Networks	N-R	0.53	0.6	0.53	0.58
	R	0.55	0.6	0.55	0.56
4. Individual CNN and LSTM Networks with SVM classifier	N-R	0.52	0.55	0.53	0.56
	R	0.49	0.6	0.5	0.56
5. Single 3-Layer CNN and LSTM Networks	N-R	0.5	0.5	0.5	0.52
	R	0.43	0.61	0.39	0.53
6. Multiple CNN's and LSTM Network	N-R	0.53	0.60	0.53	0.58
	R	0.55	0.6	0.56	**0.59**
7. Single 3-Layer CNN and bi-LSTM Network	N-R	0.52	0.59	0.53	0.57
	R	0.50	0.57	0.50	0.55
8. Multiple CNN's and bi-LSTM Network	N-R	0.54	0.60	0.55	**0.59**
	R	0.55	0.6	0.56	**0.59**

Table no 3: Comparison of the state-of-the-art approaches with the best outcomes of this report

Study	Network System	Word Embedding	Dataset (labeled Tweets)	Accuracy
Baziotis et al. [22]	bi-LSTM	GloVe	~50.000	**0.65**
Cliche [23]	CNN+LSTM	GloVe FastText Word2Vec	~50.000	**0.65**
Deriu et al. [20]	CNN	GloVe Word2Vec	~300.000	**0.65**
Rouvier and Favre [21]	CNN	Lexical, POS, Sentiment	~20.000	0.61
Wange et al. [27]	CNN+LSTM	Regional Word2Vec	~8.500	1.341[a]
Current study	CNN+LSTM	Regional, GloVe	~31.000	0.59

IV. Conclusion

The CNN and LSTM networks are utilized to assess various configurations of deep learning methods for sentimental research on Twitter data in this study. This evaluation yielded somewhat lower, but similar results when compared to state-of-the-art methods, allowing us to make valid judgments about the different programmers. The platforms' low efficiency highlighted flaws in the field of CNN and LSTM networks. The usage of a combination of CNN and LSTM networks is more efficient than using them separately in terms of setup. This is due to the effective technique of using LSTM networks to reduce the dimension of CNN while preserving word dependencies. Several CNN and LSTM networks, on the other hand, enhance device efficiency. The difference in accuracy between the different data sets shows that, as expected, a good dataset is the most important factor in increasing the effectiveness of these systems. As a result, it seems that more effort and money is spent on creating effective training sets rather than experimenting with different CNN and LSTM network variants and configurations. This paper's contribution is to enable the testing and experimentation of multiple deep neural network configurations inside a single data set and evaluation structure for two distinct word embedding frameworks, allowing them to explain their benefits and limits.

References

[1]. L. L. Bo Pang, "Opinion Mining and Sentiment Analysis Bo", Found. Trends® Inf. Retr., vol. 1, no. 2, pp. 91–231, 2008.

[2]. K. Fukushima, "Neocognition: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biol. Cybern., vol. 202, no. 36, pp. 193–202, 1980.

[3]. Y. Lecun, P. Ha, L. Bottou, Y. Bengio, S. Drive, eta R. B. Nj, "Object Recognition with Gradient-Based Learning", in Shape, contour and grouping in computer vision, Heidelberg: Springer, 1999, pp. 319–345.

[4]. D. E. Ruineihart, G. E. Hinton, eta R. J. Williams, "Learning internal representations by error propagation", 1985.

[5]. S. Hochreiter eta J. UrgenSchmidhuber, "Ltsm", Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[6]. Y. Kim, "Convolutional Neural Networks for Sentence Classification", arXivPrepr. arXiv1408.5882., 2014.

[7]. C. N. dos Santos eta M. Gatti, "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts", in Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014, pp. 69–78.

[8]. P. Ray eta A. Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis", Appl. Comput. Informatics, 2019.

[9]. S. Lai, L. Xu, K. Liu, eta J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification", Twenty-ninth AAAI Conf. Artif. Intell., pp. 2267–2273, 2015.

[10]. D. Tang, B. Qin, eta T. Liu, "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification", in Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 1422–1432.

[11]. K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, eta Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation", arXivPrepr. arXiv1406.1078, 2014.

[12]. J. Chung, C. Gulcehre, K. Cho, eta Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", arXivPrepr. arXiv1412.3555, 2014.

[13]. L. Zhang, S. Wang, eta B. Liu, "Deep Learning for Sentiment Analysis: A Survey", Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, pp. e1253, 2018.

[14]. T. Mikolov, K. Chen, G. Corrado, eta J. Dean, "Efficient Estimation of Word Representations in Vector Space", arXivPrepr. arXiv1301.3781, 2013.

[15]. J. Pennington, R. Socher, eta C. Manning, "Glove: Global Vectors for Word Representation", Proc. 2014 Conf. Empir. Methods Nat. Lang. Process., pp. 1532–1543, 2014. [16]H. Kwak, C. Lee, H. Park, eta S. Moon, , in Proceedings of the 19th international conference on World wide web, 2010, pp. 591–600.

[16]. A. Go, R. Bhayani, eta L. Huang, "Proceedings - Twitter Sentiment Classification using Distant Supervision (2009).pdf", CS224N Proj. Report, Stanford, vol. 1, no. 12, 2009.

[17]. A. Srivastava, V. Singh, eta G. S. Drall, "Sentiment Analysis of Twitter Data", in Proceedings of the Workshop on Language in Social Media, 2011, pp. 30–38.

[18]. E. Kouloumpis, T. Wilson, eta M. Johanna, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", in Fifth International AAAI conference on weblogs and social media, 2011.

[19]. J. Deriu, M. Gonzenbach, F. Uzdilli, A. Lucchi, V. De Luca, eta M. Jaggi, "SwissCheese at SemEval-2016 Task 4: Sentiment Classification Using an Ensemble of Convolutional Neural Networks with Distant Supervision", Proc. 10th Int. Work. Semant. Eval., pp. 1124–1128, 2016.

[20]. M. Rouvier eta B. Favre, "SENSEI-LIF at SemEval-2016 Task 4 : Polarity embedding fusion for robust sentiment analysis", Proc. 10th Int. Work. Semant. Eval., pp. 207–213, 2016.

[21]. C. Baziotis, N. Pelekis, eta C. Doulkeridis, "DataStories at SemEval2017 Task 4: Deep LSTM with Attention for Message-level and Topicbased Sentiment Analysis", Proc. 11th Int. Work. Semant. Eval., pp. 747–754, 2017.

[22]. M. Cliche, "BB twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs", arXivPrepr. arXiv1704.06125.

[23]. P. Bojanowski, E. Grave, A. Joulin, eta T. Mikolov, "Enriching Word Vectors with Subword Information", Trans. Assoc. Comput. Linguist., Vol 5, pp. 135–146, 2016.

[24]. T. Lei, H. Joshi, R. Barzilay, T. Jaakkola, K. Tymoshenko, A. Moschitti, eta L. Marquez, "Semi-supervised Question Retrieval with Gated Convolutions", arXiv Perry. arXiv1512.05726, 2015.

[25]. Y. Yin, S. Yangqiu, eta M. Zhang, "NNEMBs at SemEval-2017 Task 4: Neural Twitter Sentiment Classification: a Simple Ensemble Method with Different Embeddings", Proc. 11th Int. Work. Semant. Eval.